

University of Groningen

Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study

Schipper, Tijmen Michiel; Goei, Sui Lin; de Vries, Siebrich; van Veen, Klaas

Published in:
International Journal of Educational Research

DOI:
[10.1016/j.ijer.2018.01.011](https://doi.org/10.1016/j.ijer.2018.01.011)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schipper, T. M., Goei, S. L., de Vries, S., & van Veen, K. (2018). Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study. *International Journal of Educational Research*, 88, 109-120. <https://doi.org/10.1016/j.ijer.2018.01.011>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study

Tijmen Schipper^{a,c*}, Sui Lin Goei^{a,b}, Siebrich de Vries^c, and Klaas van Veen^c

^a *Department of Human Movement and Education, Windesheim University of Applied Sciences, Zwolle, the Netherlands.*

^b *LEARN! Research Institute, VU University Amsterdam, Amsterdam, the Netherlands.*

^c *Department of Teacher Education, University of Groningen, Groningen, the Netherlands.*

*Corresponding author: P.O. Box 10090, Campus 2-6, 8000 GB Zwolle, The Netherlands. E-mail address: t.schipper@windesheim.nl, telephone number: +31(0)88 46 99 911.

Abstract

Teachers are expected to address a broad range of diverse pupil needs but do not always feel capable or lack the skills to meet these high expectations. The professional development approach Lesson Study may address this. Therefore, this study examines whether participating in Lesson Study influences teachers' beliefs of self-efficacy and (adaptive) teaching behaviour. A quasi-experimental mixed methods design was used to compare pretest and posttest data of intervention and comparison group teachers ($N = 48$). Significantly different results between the two groups arise in terms of efficacy in pupil engagement as well as in classroom management and instructional behaviour. Immediate stimulated recall interviews provide insight in these outcomes and illustrate to what extent teachers addressed pupils' educational needs.

Keywords: lesson study; teacher self-efficacy; professional development; adaptive teaching; quasi-experimental.

Funding

This work was supported by The Netherlands Organization for Scientific Research (NWO) under Grant Number: 023.005.105.

1. Introduction

Responding to a variety of pupils' needs and backgrounds in order to reach the desired instructional goals, seems to comply to the demands of learning in the 21st century and its global trend towards more adaptive teaching in inclusive settings (Schleicher, 2016; UNESCO, 2009). However, teachers have difficulties providing differentiated instruction to respond appropriately to pupils' individual learning needs (Bruggink, Goei, & Koot, 2015; Randi & Corno, 2005) and lack confidence or feel unprepared for these teaching practices (Dixon, Yssel, McConnell, & Hardin, 2014; Wan, 2016). This calls for confident, self-efficacious teachers (Tschannen-Moran & Woolfolk Hoy, 2001), who are able to adapt their teaching to pupils' diverse learning needs (Summers, Davis, & Woolfolk Hoy, 2017). The professional development (PD) approach Lesson Study (LS) is believed to address these issues (Puchner & Taylor, 2006; Ylonen & Norwich, 2015). Following this rationale, this study aims to determine whether participating in LS influences beliefs of self-efficacy and adaptive teaching behaviour.

In Japan, LS (translated as *jugyou kenkyuu*) has been an integral part of teaching for more than a century (Takahashi & McDougal, 2016). After the publication of 'The Teaching Gap' (Stigler & Hiebert, 1999), this form of PD rapidly expanded around the globe (Hadfield & Jopling, 2016; Huang & Shimizu, 2016). Dudley (2015) argues that it is currently the world's fastest growing approach to teacher learning. Teachers participating in LS follow systematic cycles of collaborative studying, planning, teaching and observing so-called 'research lessons', focusing on the learning of pupils. Subsequently, the research lesson is evaluated and refined in order to improve classroom practice and pupil learning (Dudley, 2013; Sims & Walsh, 2009).

The reviews of Huang and Shimizu (2016) and Xu and Pedder (2015) show, among other things, that participating in LS contributes to teachers' beliefs of self-efficacy and the

quality of teaching and learning as well as a more explicit focus on and more awareness of the diverse learning needs of pupils. However, it turns out that “most of the research carried out into LS has adopted a small-scale, qualitative, exploratory and inductive mode of inquiry” (Xu & Pedder, 2015, p. 49). This study contributes to the limited research on LS in which (quasi-)experimental mixed methods designs have been deployed. It includes data from different perspectives using teacher questionnaires, classroom observations and interviews. This enables us to not only determine whether participating in LS leads to changes in teachers’ beliefs of self-efficacy and their (adaptive) teaching behaviour, but also allows us to determine whether a relationship between these two constructs exists and to examine possible explanations using the qualitative data. We first elaborate on LS, teacher self-efficacy and adaptive teaching behaviour.

2. Theoretical framework

2.1. Teacher PD through LS

Although there is widespread consensus about the importance of teacher PD, the actual form of PD activities varies tremendously (Kennedy, 2016) and evaluation of these programmes may serve different purposes (Merchie, Tuytens, Devos, & Vanderlinde, 2016). The often cited conceptual framework of Desimone (2009) synthesizes a large body of research on PD and points to the “interactive, nonrecursive relationships between critical features of PD, teacher knowledge and beliefs, classroom practice, and student outcomes” (p. 184). The model distinguishes five critical features of PD in order to be effective: (1) content focus, (2) active learning, (3) coherence (consistency with teachers’ knowledge and beliefs as well as with school and macro-level reforms and policies), (4) duration (span of time and number of hours spent on the PD activity), and (5) collective participation.

Lewis and Perry (2017) applied this framework to LS and show how these effective

characteristics of high quality teacher professional learning are integrated in LS. The cyclical features that LS embodies relate to: (1) a clear research purpose, (2) an in-depth investigation of lesson material, research articles, and available curricula, (3) collaborative planning of the research lesson, (4) teaching the research lesson by one teacher and live observation by the other group members, (5) a thorough post-lesson discussion, preferably guided by a 'knowledgeable other' (Takahashi & McDougal, 2016), which is often a university faculty member or someone from a professional association (Lee, 2015), and (6) dissemination of the results via publishing articles or organising open houses (Lewis, Perry, & Hurd, 2009). In the US and Europe, refining and re-teaching the research lesson are integrated in the LS cycle (Dudley, 2013; Saito & Atencio, 2013), whereas this is not common practice in Japan (Fuji, 2014).

Following the conceptual framework of Desimone (2009), the rapidly growing body of research on LS shows that participating in LS results in increased teacher knowledge and skills (e.g., Dudley, 2013; Leavy & Hourigan, 2016; Lewis, Perry, & Hurd, 2009; Takahashi & McDougal, 2016; Vrikki, Warwick, Vermunt, Mercer, & Van Halem, 2017), as well as changes in attitudes and beliefs (e.g., Cajkler, Wood, Norton, & Pedder, 2014; Puchner & Taylor, 2006; Schipper, Goei, De Vries, Van Veen, 2017; Sibbald, 2009). This, in turn, leads to changes in instructional practice (e.g., Lewis, Perry, & Murata, 2006; Lewis, Perry, & Hurd, 2009) and improved pupil learning (e.g., Norwich & Ylonen, 2013; Lewis & Perry, 2017). Lee (2015) argues that in the context of LS, teacher knowledge refers to subject matter knowledge, knowledge of instruction, the capacity to observe pupils, and the connection of daily practice to long-term goals.

The systematic approach of LS allows teachers to devote considerable thought to predicting how pupils might react in different situations (Dudley, 2013) and how they would address pupils' learning needs (Sims & Walsh, 2009; Van Halem, Goei, & Akkerman, 2016;

Ylonen & Norwich, 2015). This explicit focus on pupil learning is repeatedly stressed in the included studies of the review by Xu and Pedder (2015). LS enables teachers to “develop the eyes to see children and how they respond and learn during research lessons” (Lee, 2015, p. 103). In the United Kingdom (UK), this focus is promoted by using ‘case pupils’ who represent different attainment groupings (Dudley, 2013). In the Netherlands, where this study took place, focusing explicitly on different educational needs is what we derive from this model by (Goei, Norwich, & Dudley, *in press*).

Despite a growing body of research showing a clear link between participating in LS and its effect on teachers’ knowledge, skills, attitudes and behaviour as well as pupil learning, only limited LS research relates specifically to teachers’ beliefs of self-efficacy (e.g., Puchner & Taylor, 2006; Sibbald, 2009).

2.2. Teacher self-efficacy in the context of (adaptive) teaching

Feelings of competence are often referred to as self-efficacy (Tschannen-Moran & Woolfolk Hoy, 2001). Bandura (1977) first described this notion as one’s beliefs or convictions to successfully execute a given type of performance. He later redefined this definition as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performance” (Bandura, 1986, p. 391). Since then, studies on self-efficacy “have been popping up like daisies in a spring field” (Zee & Koomen, 2016, p. 981), illustrating the popularity of this construct. In the context of education, self-efficacy is often referred to as teacher self-efficacy (TSE) and is defined as “teachers’ belief or conviction that they can influence how well students learn, even those who may be difficult or unmotivated” (Guskey & Passaro, 1994, p. 628). One of its claims is that TSE might be a vital predictor of teacher behaviour (Summers, Davis, & Woolfolk Hoy, 2017; Tschannen-Moran & Woolfolk Hoy, 2001). The underlying general assumption of TSE is that when teachers feel more

confident to meet pupils' instructional needs, they tend to focus more on improving their teaching activities (Summers, Davis, & Woolfolk Hoy, 2017).

It is argued that teachers with a higher sense of self-efficacy are associated with a higher quality of classroom environment as a result of processes that relate to pupil involvement, instructional strategies, and classroom management (Tschannen-Moran & Woolfolk Hoy, 2001; Zee & Koomen, 2016). In addition, they seem to be more likely to use instructional knowledge and skills that they have gained in PD activities (Zee & Koomen, 2016). These higher levels of TSE may increase pupils' achievement and motivation (Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). The contrary applies to teachers with a lower sense of self-efficacy (Skaalvik & Skaalvik, 2007). These teachers "may ask easier questions, allow less time for answering, and give fewer prompts, and express less warmth in their interactions with students" (Summers, Davis, & Woolfolk Hoy, 2017, p. 19). To summarise, the popularity of TSE may be caused by its cyclical nature which can be described as "stronger self-efficacy beliefs are believed to result in greater efforts by teachers, which in turn leads to better performances, which again provides information for forming higher efficacy evaluations" (Malinen et al., 2013, p. 35).

In the context of adaptive teaching practices, the review of Zee and Koomen (2016) reports that self-efficacious teachers have been shown to be more positive toward inclusive education practices. Teachers with higher levels of self-efficacy are claimed to be more willing to address their pupils' learning needs (Dixon, Yssel, McConnell, & Hardin, 2014) and Suprayogi, Valcke, and Godwin (2017), for example, report a significant association between higher levels of TSE and higher levels of experimenting with differentiated instruction practices as well as a more positive attitude towards differentiated instruction. In addition, teachers who spend more hours on PD specifically focusing on differentiation seem to feel more efficacious in differentiating instruction in their classes (Dixon, Yssel,

McConnell, & Hardin, 2014).

In terms of measuring TSE, various instruments have been developed in the previous decades (Zee & Koomen, 2016). One frequently used instrument, the Teachers' Sense of Efficacy Scale (TSES) (Tschannen-Moran & Woolfolk Hoy, 2001), is believed to be superior to other measures of TSE "in that it has a unified and stable factor structure and assesses a broad range of capabilities that teachers consider important to good teaching" (Woolfolk Hoy & Burke Spero, 2005, p. 354). The scale treats TSE as a task-specific, three-dimensional construct (Zee & Koomen, 2016) addressing pupil engagement, instructional strategies and classroom management.

2.3.Adaptive teaching

Drawing on previous research, adaptive teaching can be described in terms of teachers' adjustments of their planning and teaching to meet the individual educational needs in order to reach the desired goals and is related to teachers' subject knowledge, familiarity with and diagnosis of pupil learning, teaching methods and classroom management (Beltramo, 2017; Brühwiler & Blatchford, 2011; Corno, 2008; Randi & Corno, 2005; Snow, 1997). Despite claims that differentiated instruction might be a newer and more detailed concept than adaptive teaching (Smit & Humpert, 2012), the literature is ambiguous about the use of both concepts and their definitions often seem to overlap (Suprayogi, Valcke, & Godwin, 2017).

Differentiated instruction often refers to actual teaching behaviour in classroom settings and can be defined as "an instructional approach that accommodates the diversity of students by (1) coping with student diversity; (2) adopting specific teaching strategies; (3) invoking a variety in learning activities; (4) monitoring individual student needs; and (5) pursuing optimal learning outcomes" (Suprayogi, Valcke, & Godwin, 2017, p. 292). What could arguably distinguish adaptive teaching from differentiated teaching, is that adaptive teaching not only refers to instructional practice, but is often described as a way thinking

about teaching and learning that also takes place outside the classroom (Corno, 2008; Randi & Corno, 2005). In that sense, adaptive teaching explicitly entails a certain mindset or vision that enables teachers to view learner variation as an opportunity rather than as an obstacle (Beltramo, 2017; Corno, 2008). Brühwiler and Blatchfrod (2011), for example, refer to adaptive teaching competency and illustrate how actual planning and teaching performances are connected to beliefs and values, motivation and self-regulation. However, this way of thinking, a philosophy, is also referred to in research that uses the term differentiated instruction (Dixon, Yssel, McConnell, & Hardin, 2014, Suprayogi, Valcke, & Godwin, 2017; Tomlinson, 2005). Given the existence of different ‘labels’ that share more or less the same ideas, i.e. to cope with and to address pupils’ different educational needs (Suprayogi, Valcke, & Godwin, 2017), and without a clear distinction between these labels, we believe that the term adaptive teaching better captures this mindset.

Although it is argued that contemporary teachers are expected to be learning-oriented adaptive experts who are able to teach increasingly diverse learners (De Vries, Jansen, Helms-Lorenz, and Van de Grift, 2015; Wan, 2016), addressing pupils’ different educational needs turns out to be difficult (Randi & Corno, 2005; Van de Grift, Helms-Lorenz, & Maulana, 2014; Van der Lans, Van de Grift, & Van Veen, 2017) and is often attributed to expert teachers (Berliner, 2001). In the Dutch context of this study, it turns out that less than half of the secondary school teachers differentiate their instruction between pupils (Dutch Inspectorate of Education, 2017; OECD, 2016).

To measure the quality of (adaptive) teaching behaviour, one prominent method is to evaluate teachers’ impact on pupils’ test scores (Chetty, Friedman, & Rockoff, 2012). A major critique regarding this value-added paradigm is that it offers only limited information about (adaptive) teacher behaviour and does not examine how PD might develop teacher behaviour (Pianta & Hamre, 2009). Observation instruments, on the contrary, have gained

more attention in the literature on teacher learning and it is repeatedly stressed that these instruments are relatively objective evaluation tools for measuring the quality of teacher behaviour (Desimone, 2009; De Vries, Helms-Lorenz, & Van de Grift, 2015; Hill, Charalambous, & Kraft, 2012; Pianta & Hamre, 2009). It is argued that these instruments provide a standardized way of measuring teacher behaviour and determining whether PD programmes actually improve teacher quality (Pianta & Hamre, 2009). As a result, various observation instruments have been developed (Patrick & Mantzicopoulos, 2016). In the Netherlands, the International Comparative Analysis of Learning and Teaching (ICALT) (Van de Grift, 2007) is widely used and has proven its strengths in terms of investigating the effectiveness of PD programmes (Van der Lans, Van de Grift, & Van Veen, 2017). The ICALT consists of six domains of teacher behaviour including one domain that explicitly focuses on adaptive teaching. Years of examining the ICALT results reveal that the domain of adaptive teaching belongs to one of the most difficult teaching skills (Van de Grift, Helms-Lorenz, & Maulana, 2014). Due to its explicit focus on pupil learning, LS may play a pivotal role in enhancing this complex skill by promoting teacher awareness of pupils' different educational needs as well as actual adaptive teaching behaviour (Schipper, Goei, De Vries, & Van Veen, 2017).

3. Present study

This study deploys a quasi-experimental mixed methods design to determine whether participating in the PD approach LS influences teachers' beliefs of self-efficacy and their adaptive teaching behaviour. The following research questions can be distinguished:

- (1) To what extent does participating in LS influence TSE?

- (2) To what extent does participating in LS influence adaptive teaching behaviour?
 - a. From the observers' perspective
 - b. From the teachers' perspective
- (3) What is the relationship between TSE and adaptive teaching behaviour?

4. Research method

4.1. Context and participants

Eight LS teams from eight different schools in the Western and Northern region of the Netherlands served as the intervention group in this study. The teams participated in one of three LS research projects executed by three Dutch universities. Two of these LS projects were funded by the Dutch Ministry of Education, Culture and Science, and cover seven of the eight included secondary schools.

Teachers were informed about the research objectives and data collection procedure via e-mail. Participation was voluntary and teachers were only included if they participated in at least two LS cycles during the academic year 2015-2016. The teaching subjects were clustered into three core academic subject clusters (languages, social sciences and sciences). In addition to the intervention group, teachers from the same subject clusters within the same schools, who were not involved in any LS activity, were invited to take part in the comparison group. Given the fact that Dutch secondary education teachers are entitled to spend 10% of their annual working hours on PD (Dutch Counsel for Secondary Education, 2016), these teachers participated in other PD activities such as attending workshops and conferences, (teacher) training courses, and reading research literature. Written informed consent was obtained from all teachers.

In total, 48 secondary school teachers ($N = 48$) were included in this study based on the criteria that the intervention group teachers conducted the intervention as intended and

that all teachers in both groups took part in the full range of instruments during both the pre- and posttest. The teachers were almost equally divided between the intervention group ($n = 26$) and comparison group ($n = 22$). Following the sample descriptions in Table 1, we conducted independent T-tests to control for baseline differences in teacher characteristics between both groups. This yielded no significant differences from which we can assume that both groups are comparable.

<INSERT TABLE 1 HERE>

4.2. The intervention

All LS teams followed the LS cycles as intended and selected objectives that relate to addressing pupils' different educational needs. In some cases this theme was accompanied by other elements such as data driven education or subject specific elements. Two of the subsidised LS projects (research project #1 and #2, Table 2) formalised this overarching objective in their research proposals. In the remaining school, the focus on addressing pupils' different educational needs in the LS team was confirmed by the school board and became evident by examining the lesson plans.

Variations between schools arose in terms of time allocation, the use of case pupils, whether LS teams were content specific or interdisciplinary, and whether an external or internal LS facilitator was involved (Table 2). The majority of teachers ($n = 19$) used case pupils and were guided by an external and trained LS facilitator. School #8 used external as well as internal LS facilitators. In this study, internal LS facilitators are defined as teachers from the same school with generally limited knowledge of and experience with LS.

The allocated time to participate in LS was unequally distributed. Teachers in schools #6 and #7 were structurally facilitated with a fixed afternoon each week to participate in LS activities, whereas in other schools meetings were planned throughout the year. In schools #1 to #5, this resulted in six meetings per LS cycle of approximately two hours each,

supplemented with an introduction meeting at the start and an overall reflection meeting at the end of two LS cycles. In school #8 this resulted in five fixed meetings of four hours per cycle, supplemented with extra time to spend on preparing the research lesson.

<INSERT TABLE 2 HERE>

4.3. Research design and data collection procedure

A repeated measures design was used. Data were collected before and after the intervention took place at the beginning and the end of one academic year (2015- 2016). Figure 1 presents an overview of the three research instruments used in this study.

<INSERT FIGURE 1 HERE>

TSE questionnaire

In order to measure levels of TSE, a Dutch online version (Goei & Schipper, 2016) of the long form of the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Woolfolk Hoy, 2001) was used. The questionnaire consists of 24 items equally divided over three subscales. Internal consistency was determined using Cronbach's alfa, indicating reliable scores: (a) efficacy in pupil engagement' ($\alpha = .73$, e.g., 'How much can you do to help pupils think critically?'), (b) efficacy in instructional strategies ($\alpha = .80$, e.g., 'How well can you respond to difficult questions from your pupils?'), and (c) efficacy in classroom management' ($\alpha = .91$, e.g., 'How much can you do to get children follow classroom rules?'). The presented Cronbach's alpha scores of the subscales were obtained in the posttest and all the items were measured on a nine-point Likert scale ranging from 1 ('nothing') to 9 ('a great deal').

The online questionnaire was distributed using the questionnaire services of Formdesk. An online version of the questionnaire was initially sent to 108 teachers of whom 63 teachers were engaged in LS and 45 teachers who were not. Two reminders were sent in order to increase the response rate. The pretest yielded 85 valid responses (response rate of

78.7%) whereas the posttest resulted in 71 valid responses (response rate of 65.7%).

Eventually, 64 teachers who participated in the pre- and posttest (treatment group: $n = 44$; comparison group: $n = 20$) could be matched using a unique personal code. These 64 teachers constitute the sample of this study.

ICALT observation instrument

Teacher behaviour was measured with the ICALT observation instrument (Van de Grift, 2007). If teachers agreed to participate, they were asked to teach a 'business as usual lesson' in order to provide representative data. The ICALT contains 32 items divided into six domains of teacher behaviour and one domain of pupil involvement. All subscales indicate reliable Cronbach's alpha scores: (1) 'Safe and stimulating learning climate' (4 items, $\alpha = .85$), (2) 'Efficient classroom management' (4 items, $\alpha = .80$), (3) 'Clarity of instruction' (7 items, $\alpha = .91$), (4) 'Activating learning' (7 items, $\alpha = .83$), (5) 'Adaptive teaching' (4 items, $\alpha = .79$), and (6) 'Learning strategies' (6 items, $\alpha = .86$). The seventh domain focuses on 'pupil involvement' and consists of three items ($\alpha = .93$). All 35 items were measured on a 4-point Likert scale ranging from 1 ('predominantly weak') to 4 ('predominantly strong'). In accordance with the TSE questionnaire, the internal consistency scores were obtained in the posttest of this study.

Following the procedure of Van de Grift, Helms-Lorenz, and Maulana (2014), the observations were carried out by observers who were specially trained to conduct the observations. The training consists of an explanation of the instrument, group discussions, and observing three video-recorded lessons. Candidates who met the consensus norm of .70 or higher were judged as eligible to participate as observer in this study. As a result, the first author conducted 45 paired observations in the pre- and posttest lessons and a fellow teacher researcher conducted the remaining three observations in the pretest and posttest lessons. Both observers had at least nine years of teaching experience in different schools and were

used to observe fellow teachers in their own practice (through co-teaching, research activities and coaching of new colleagues). The first author was aware of how the research sample was composed and knew, as a consequence, which teacher was part of the intervention or comparison group. The fellow teacher researcher did not have this information.

Immediate stimulated recall interviews

Short stimulated recall interviews were conducted to measure teachers' perspectives in terms of their (adaptive) teacher behaviour in the pretest and posttest lessons. The idea behind stimulated recall is that participants verbalise their thoughts and actions concerning a specific teaching situation (Beers, Boshuizen, Kirschner, Gijssels, & Westendorp, 2008; Vesterinen, Toom, & Patrikainen, 2010).

The interviews were conducted by the same researchers who observed the lesson with the ICALT instrument (i.e. 45 teachers were interviewed by the first author of this study and three teachers were interviewed by a fellow teacher researcher) and took place directly after the observed lesson or within three hours after the observed lesson in case teachers' schedules did not allow to meet directly after the lesson. The interviews lasted approximately ten minutes in the pretest and fifteen minutes in the posttest, and consisted of twelve and twenty questions respectively. Both interviews focused on teachers' thoughts and actions during the lessons as well as several questions that focused on the intended goals of the lessons and whether unexpected events arose. Several specific questions referring to adaptive teaching were included in the interviews (e.g., 'To what extent were you able to cope with pupils' different educational needs in this lesson?'). The additional posttest questions focused on teacher development practices (e.g., 'Did you show particular teaching behaviour in this lesson that you have worked on during this academic year?'). Furthermore, questions focusing on the PD activities were posed (e.g., 'Have you gained new knowledge in the past academic year?'). Given the fact that the researchers who interviewed the teachers also observed the

lessons, they could easily relate to particular teaching situations which caught their eye or which were discussed by the teachers.

4.4.Data analysis

Although the data collection using the different research instruments had no particular order, the analysis of the data, however, had a more sequential character in line with an explanatory sequential mixed methods design (Creswell, 2012). In essence, the quantitative data were used as a starting point of the analysis aiming to detect patterns, followed by an analysis of the qualitative data in order to gain deeper insights into these patterns and to focus on adaptive teaching behaviour in particular. The second aim of the research design was to increase internal validity by combining two perspectives focusing on teacher behaviour: teachers' perspective on their own teacher behaviour as well as an independent observers' perspective on the same teacher behaviour.

Quantitative analysis

In terms of baseline differences between both groups, independent t-tests for the subscales in the TSE and ICALT instruments did not yield any significant differences between both groups. Based on these results we can assume that during the pretest of this study, the intervention group and comparison group are comparable in terms of TSE and teacher behaviour.

Furthermore, testing the subscales of the TSE instrument for normality using Kolmogorov-Smirnov tests did not yield significant deviations from a normal distribution in both groups. Applying the same tests for the subscales of the ICALT results in a non-normal distribution in two of the seven subscales in the intervention group (adaptive teaching $D(21) = .28, p < .01$ and pupil involvement $D(21) = .20, p < .05$), and one of the seven subscales in the comparison group (teaching learning strategies $D(18) = .23, p < .05$). Since the majority of the

subscales for both groups in both instruments are normally distributed and the assumption of sphericity was not violated, we decided to conduct parametric repeated-measures ANOVA tests. This test has shown to be robust to violations of its assumptions (Field, 2013).

Qualitative analysis

Following the quantitative analysis of the TSE and ICALT data, the analysis of the qualitative immediate stimulated recall interviews consisted of the following procedure: the pretest and posttest interview annotations were thoroughly read. Subsequently, all answers that revealed a particular growth in teacher behaviour in the posttest lesson were highlighted and related to their PD activities in the particular academic year. A third step was to compare differences between teachers in the intervention group and the comparison group. Lastly, three categories were constructed for each teacher to express the extent to which teachers indicated growth in their teaching behaviour with a specific focus on adaptive teaching ('no growth', 'reasonable growth' or 'substantial growth'), and whether this growth was related to LS or other PD activities.

5. Results

5.1.Changes in teachers' sense of self-efficacy

Table 3 shows the mean and standard deviation values for both groups in terms of levels of TSE in the pretest and posttest. Paired t-tests for the subscales within each group report a significant increase for the subscale 'efficacy in instructional strategies' in the intervention group ($t(25) = -2.64, p < .05$).

<INSERT TABLE 3 HERE>

Testing for differences between both groups using a repeated measures ANOVA yields a significant result in the subscale ‘efficacy in pupil engagement’ ($F(1.00, 46.00) = 4.51, p < .05$). The different patterns are illustrated in figure 2.

<INSERT FIGURE 2 HERE>

5.2.Changes in adaptive teacher behaviour

From the observer perspective using ICALT

The mean scores and standard deviations for both groups in the pretest and posttest are presented in Table 4. Paired t-tests¹ on the subscale level within each group result in several significant increases in the intervention group. This applies to the subscales (3) ‘clarity of instruction’ ($t(25) = -2.28, p < .05$), (4) ‘activating learning’ ($t(25) = -2.93, p < .01$), and (6) ‘teaching learning strategies’ ($t(25) = -2.52, p < .05$). In addition to this, the overall Rash score (Van der Lans, Van de Grift, & Van Veen, 2017) in the intervention group increases significantly as well ($t(25) = -2.83, p < .01$). In the comparison group, two significant outcomes arise. The subscale (2) ‘efficient classroom management’ shows a significant decrease in the posttest ($t(21) = 2.11, p < .05$), whereas the subscale (5) ‘adaptive teaching’ shows a significant increase in the posttest ($t(21) = -2.82, p < .05$).

<INSERT TABLE 4 HERE>

Running a repeated measures ANOVA to determine differences between both groups yields significant outcomes for the subscale ‘efficient classroom management’ ($F(1.00, 46.00) = 7.71, p < .05$) and for the subscale ‘clarity of instruction’ ($F(1.00, 46.00) = 6.62, p < .05$).

¹ Given the non-normal distribution of several ICALT subscales, the results are checked using non-parametric Wilcoxon signed rank tests. This yield the same significant outcomes.

Figure 3 shows the patterns of the two domains which yield significantly different results between both groups over time ('classroom management' and 'clarity of instruction'), as well as the domain 'adaptive teaching' which is central in this study but does not yield a significant difference between groups.

<INSERT FIGURE 3 HERE>

5.3. Teachers' perspectives in terms of their adaptive teacher behaviour

The majority of teachers in the intervention group ($n = 22$) report professional growth in the posttest interviews. This growth can be divided in a substantial growth ($n = 13$) and a reasonable growth ($n = 9$). The remaining teachers who do not report any growth ($n = 4$) argue that they did not perform differently in the posttest. Sixteen of the twenty-six teachers in the intervention group specifically refer to growth in adaptive teaching in the posttest interviews, of which seven teachers argue that this growth must have visible in the observed posttest lesson. Teachers in the intervention group who report changes in their adaptive teacher behaviour, mention that they:

- pay more attention to differences between pupils using available pupil information;
- ask pupils to formulate their own learning goals for the lesson;
- use learning tasks in which pupils could choose what they would work on;
- use a model to classify pupils in different groups based on their learning needs and consequently prepare their lessons;
- use differentiated lesson material or instruction strategies to address the different cognitive levels and learning preferences of pupils;
- explicitly experiment with adaptive teaching material in order to make lessons more meaningful for all pupils;

- address the more quiet as well as the more high achieving pupils.

Deeper analysis of the data reveals that teachers who report LS as a driver of this growth highly appreciate the intensive collaboration with colleagues and the focus on ‘case pupils’. In addition to this, active experimenting in teachers’ daily practice seems to be one of the elements that LS promoted.

A different image arises in the comparison group. Although the majority of teachers ($n = 16$) indicate professional growth, in most cases this is indicated as reasonable growth ($n = 13$) and in fewer cases as substantial growth ($n = 3$). Furthermore, only two teachers in the comparison group specifically refer to adaptive teaching. Teachers in this group obviously do not refer to LS as a driver of their professional growth, but mention different PD activities in which they participated such as exchanging information with colleagues, finishing teacher training programmes, reading about certain topics or participating in a course.

The results of three teachers in the comparison group might have been biased due to their professional activities. One teacher participated in a four-day PD course focusing specifically on constructing differentiated lessons. Two teachers argue that they exchanged experiences with several colleagues in their schools who were part of a LS team. This motivated them to pay more attention to differentiated instruction in their own lessons. Since these teachers were not participating actively in LS and exchanging or reading about adaptive teaching is common practice in the teaching profession, they remained part of the comparison group.

5.4.Relating teachers’ sense of self-efficacy to (adaptive) teaching behaviour

In this section we relate TSE to adaptive teaching behaviour. Pearson correlation coefficients were calculated to assess the relationship between the TSE subscales and the ICALT domains (Table 5). The correlations within both instruments are highly significant assuming a strong

internal validity. The subscales of both instruments are significantly correlated in terms of the TSE subscale ‘efficacy in classroom management’ and most of the ICALT subscales, including ‘adaptive teaching’. The TSE subscale ‘efficacy in pupil engagement’ is significantly correlated with the ICALT subscales ‘activating learning’ and ‘adaptive teaching’. The results show no significant relationship between the subscale ‘efficacy in instructional strategies’ and any of the ICALT subscales.

<INSERT TABLE 5 HERE>

6. Conclusions and discussion

This research examined to what extent the PD approach LS contributes to more self-efficacious feelings of teachers as well as their actual teaching behaviour focusing specifically on adaptive teaching. The use of a quasi-experimental mixed methods design enabled us to detect significant differences between teachers who participated in at least two LS cycles over the course of one academic year (intervention group) as opposed to teachers who were part of a comparison group. Focusing on the first research question, to what extent participating in LS influenced TSE, a significant difference between both groups, in favour of the intervention group, was found in terms of ‘efficacy in pupil engagement’. Based on this result, we argue that LS contributes to feelings of competence necessary to include all pupils. Items that are part of this subscale refer to paying explicit attention to pupils who are less motivated or show low interest as well as supporting all pupils to think critically and to make them value learning. The TSE findings are particularly of interest given the fact that efficacy beliefs tend to remain quite stable, especially for experienced teachers (Tschannen-Moran & Woolfolk Hoy, 2001).

Applying the same statistical procedures for answering the second research question, to what extent participating in LS influences adaptive teaching behaviour, we found no significant differences in observed behaviour between both groups in terms of the subscale

‘adaptive teaching’. However, we did find a significant difference for the subscales ‘efficient classroom management’ and ‘clarity of instruction’. Again, the increased values over the course of one academic year were in favour of the intervention group. This may indicate that LS contributes to teacher behaviour that focuses on ensuring effective classroom management as well as delivering clear classroom instruction. These domains of teaching behaviour can be interpreted as essential, yet do not belong to the more complex teaching domains such as adaptive teaching (Van der Lans, Van de Grift, & Van Veen, 2017). What is relevant in this context, is that the items in the subscale ‘clear classroom instruction’ contain elements that refer to involving pupils in the lesson and stimulating pupils to perform as best as they can. This corroborates the TSE findings relating to the subscale ‘efficacy in pupil engagement’.

From the perspective of the teachers, the immediate stimulated recall interviews explored whether teachers experienced professional growth over the course of the academic year and to what extent LS contributed to this. The results indicate that the intervention group teachers paid more attention to differences between pupils as a result of experimenting with differentiated lesson material and instructional strategies in order to make learning more meaningful for all pupils. LS seems to enable this. On the contrary, the comparison group teachers indicate professional growth in fewer cases or refer to professional growth that they gained as a result of other PD activities.

In terms of the third research question that focuses on the relationship between TSE and adaptive teacher behaviour, the significant correlations between TSE subscales and several teaching behaviour domains provide support for a potential link between teachers’ sense of self-efficacy and effective teaching behaviour. In particular, the TSE subscales ‘efficacy in classroom management’ and ‘efficacy in pupil engagement’ are both significantly correlated with adaptive teaching behaviour. Although we cannot make statements in terms of causality, the results may indicate that if teachers become more self-efficacious in terms of

classroom management and pupil engagement, that this affects the way teachers address pupils different educational needs.

No significant relationship was found between ‘efficacy in instructional strategies’ and any of the ICALT teaching domains. This is remarkable given that this subscale contains elements of specific (adaptive) teacher behaviour such as responding to difficult questions of pupils or adjusting the lesson to the proper level for individual pupils. It is unclear why this TSE subscale deviates from the other TSE subscales in terms of a significant relationship with one or more ICALT domains.

Despite these promising results there are several limitations in this study. Firstly, although the richness of the data deriving from the different instruments in this to measure the influence of LS on TSE and (adaptive) teaching behaviour, the research sample is relatively small. A bigger sample could have led to even clearer patterns and differences between both groups.

Secondly, although this study reports significant differences and the immediate stimulated recall interviews indicate professional growth in adaptive teaching, no significant differences were found in terms of adaptive teaching behaviour using the ICALT instrument. Given the clear focus on adaptive teaching in all LS teams, we assumed that, if a significant difference would arise, it would have been in this domain. A potential reason for not finding a significant effect on this domain could be related to the complexity of adaptive teaching (Van der Lans, Van de Grift, & Van Veen, 2017). Following this rationale, it could simply mean that teachers need more time and more support in LS teams in order to detect observable effects. Another potential reason is that teachers who participated in LS report more awareness of pupils’ different educational needs, but this does not necessarily result in structural changes in adaptive teaching behaviour. This would corroborate the findings of our earlier study (Schipper, Goei, De Vries, & Van Veen, 2017). A third potential reason could be

related to the way we measure adaptive teaching behaviour. The observation instrument measures observable behaviour during two fixed moment, but does not capture subtle remarks, compliments and cues in which teachers address pupils' different educational needs based on knowledge that teachers have gained about their pupils (in LS). In the immediate stimulated recall interviews, teachers were asked these types of questions. This could explain the different results in both instruments relating to adaptive teaching.

A third limitation in this study is that teachers were followed over the course of one academic year which may not be sufficient to yield structural changes in teacher behaviour (Desimone, 2009). A longitudinal design could address these issues. This also applies to the teachers who were part of the comparison group. Their PD activities varied a lot and it would be interesting to see how their participation in PD activities developed over time and to what extent this would influence their TSE and adaptive teaching competence.

7. Concluding remarks

This study belongs to one of the few studies that uses a quasi-experimental mixed methods design to examine the effects of participating in LS on teachers' feelings of self-efficacy and their adaptive teaching behaviour. We draw attention to the complementary perspectives of self-reports, via questionnaires and immediate stimulated recall interviews, as well classroom observation, using a well-validated observation instrument. Based on this multi-perspective approach and its outcomes, we argue that LS is a powerful PD approach that enables teachers to become more self-efficacious and competent in certain teaching behaviour. We believe that the explicit focus on pupil learning and the unique opportunities to collaboratively experiment with new teaching elements play a vital role in enabling this. However, more research is needed to strengthen this research base in which (quasi-)experimental designs are used to examine the effects of LS as well as explicating the underlying mechanisms that make LS effective. With this new knowledge, LS practice could be further improved.

Acknowledgements

The authors would like to express their thanks to prof. dr. Wouter van Joolingen (Utrecht University) and dr. Cor Suhre (University of Groningen) for their advice in analysing the quantitative data.

References

- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioural change. *Psychological Review*, 84(2), 191-215.
- Bandura, A. (1986). *Social foundations of thought and action. A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., Gijssels, W., & Westendorp, J. (2008). Cognitive load measurements and stimulated recall interviews for studying the effects of information and communications technology. *Educational Technology Research and Development*, 56(3), 309-328.
- Beltramo, J. L. (2017). Developing adaptive teaching practices through participation in cogenerative dialogues. *Teaching and Teacher Education*, 63, 326-337.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463-482.
- Bruggink, M., Goei, S. L., & Koot, H. M. (2015). Teachers' capacities to meet students' additional support needs in mainstream primary education. *Teachers and Teaching*, 22(4), 448-460, doi: 10.1080/13540602.2015.1082727.
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95-108.
- Cajkler, W., Wood, P., Norton, J., & Pedder, D. (2014). Lesson study as a vehicle for collaborative teacher learning in a secondary school. *Professional Development in Education*, 40(4), 511-529.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2012). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. (NBER Working Paper 17699). Cambridge, MA: National Bureau of Economic Research.
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161-173.

- Creswell, J. W. (2012). *Educational Research. Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualization and measures. *Educational Researcher*, 38(3), 181-199.
- De Vries, S., Jansen, E. P. W. A., Helms-Lorenz, M., & Van de Grift, W. J. C. M. (2015). Student teachers' participation in learning activities and effective teaching behaviours. *European Journal of Teacher Education*, 38(4), 460-483.
- Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, 37(2), 111-127.
- Dudley, P. (2013). Teacher learning in Lesson Study: What interaction-level discourse analysis revealed about how teachers utilised imagination, tacit knowledge of teaching and fresh evidence of pupils learning, to develop practice knowledge and so enhance their pupils' learning. *Teaching and Teacher Education*, 34, 107-121.
- Dudley, P. (2015). How Lesson Study works and why it creates excellent learning and teaching. In P. Dudley (Ed.), *Lesson Study. Professional learning for our time*. New York: Routledge.
- Dutch Counsel for Secondary Education (2016). *CAO VO 2016/2017. Collectieve arbeidsovereenkomst voor het voortgezet onderwijs*. [Collective Labour Agreement for secondary education 2016/2017]. Utrecht, The Netherlands: Dutch Counsel for Secondary Education.
- Dutch Inspectorate of Education (2017). *De staat van het onderwijs. Onderwijsverslag 2015/2016*. [The state of Education. Educational report 2015/2016]. Utrecht, The Netherlands: Dutch Inspectorate of Education.

- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). London: Sage.
- Fuji, T. (2014). Implementing Japanese Lesson Study in foreign countries: Misconceptions revealed. *Mathematics Teacher Education and Development*, 16(1), 65-83.
- Goei, S. L., Norwich, B., & Dudley, P. (in press). *Lesson Study for Inclusive Teaching*. Singapore: Routledge.
- Goei, S. L., & Schipper, T. M. (2016). *Teachers' Sense of Self Efficacy Scale: Teachers' opinions. Experimental version 2.0*. Amsterdam: VU University.
- Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31(3), 627-643.
- Hadfield, M., & Jopling, M. (2016). Problematizing lesson study and its impacts: Studying a highly contextualised approach to professional learning. *Teaching and Teacher Education*, 60, 203-214.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64.
- Huang, R., & Shimizu, Y. (2016). Improving teaching, developing teachers and teacher educators, and linking theory and practice through lesson study in mathematics: An international perspective. *ZDM Mathematics Education*, 48(4), 393-409.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945-980.
- Leavy, A. M., & Hourigan, M. (2016). Using lesson study to support knowledge development in initial teacher education: insights from early number classrooms. *Teaching and Teacher Education*, 57, 161-175.
- Lee, C. K. E. (2015). Examining education rounds through the lens of lesson study. *International Journal of Educational Research*, 73, 100-106.

- Lewis, C., & Perry, R. (2017). Lesson Study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, 48(3), 261-299.
- Lewis, C. C., Perry, R. R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: a theoretical model and North American case. *Journal of Mathematics Teacher Education*, 12, 285-304.
- Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher*, 35(3), 3-14.
- Malinen, O., Savolainen, H., Engelbrecht, P., Xu, J., Nel, M., Nel, N., & Tlale, D. (2013). Exploring teacher self-efficacy for inclusive practices in three diverse countries. *Teaching and Teacher Education*, 33, 34-44.
- Merchie, E., Tuytens, M., Devos, G., & Vanderlinde, R. (2016). Evaluating teachers' professional development initiatives: towards an extended evaluative framework. *Research Papers in Education*, DOI: 10.1080/02671522.2016.1271003.
- Norwich, B. & Ylonen, A. (2013). Design based research to develop the teaching of pupils with moderate learning difficulties (MLD): Evaluating lesson study in terms of pupil, teacher and school outcomes. *Teaching and Teacher Education*, 34, 162-173.
- OECD (2016). *Netherlands 2016: Foundations for the future*. Reviews of National Policies for Education. Paris, France: OECD Publishing.
<http://dx.doi.org/10.1787/9789264257658-en>
- Patrick, H., & Mantzicopoulos, P. (2016). Is effective teaching stable? *The Journal of Experimental Education*, 84(1), 23-47.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119.

- Puchner, L. D., & Taylor, A. R. (2006). Lesson study, collaboration and teacher efficacy: Stories from two school-based math lesson study groups. *Teaching and Teacher Education*, 22, 922-934.
- Randi, J., & Corno, L. (2005). Teaching and learner variation. *British Journal of Educational Psychology*, 2(3), 47-69.
- Saito, E., & Atencio, M. (2013). A conceptual discussion of lesson study from a micro-political perspective: Implications for teacher development and pupil learning. *Teaching and Teacher Education*, 31, 87-95.
- Schipper, T. M., Goei, S. L., De Vries, S., & Van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of Lesson Study. *Teaching and Teacher Education*, 68, 289-303.
- Schleicher, A. (2016). *Teaching excellence through professional learning and policy reform: Lessons from around the world*. International Summit on the Teacher Profession, OECD Publishing, Paris.
- Sibbald, T. (2009). The relationship between Lesson Study and Self-Efficacy. *School Science and Mathematics*, 109(8), 450-460.
- Sims, L., & Walsh, D. (2009). Lesson Study with preservice teachers: Lessons from lessons. *Teaching and Teacher Education*, 25, 724-733.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, 99(3), 611-625.
- Smit, R., & Humpert, W. (2012). Differentiated instruction in small schools. *Teaching and Teacher Education*, 28(8), 1152-1162.
- Snow, R. E. (1997). Aptitudes and symbol systems in adaptive classroom teaching. *The Phi Delta Kappan*, 78(5), 354-360.

- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Summers, J. J., Davis, H. A., & Woolfolk Hoy, A. (2017). The effects of teachers' efficacy beliefs on pupils' perceptions of teacher relationship quality. *Learning and Individual Differences, 53*, 17-25.
- Suprayogi, M. N., Valcke, M., & Godwin, R. (2017). Teachers and their implementation of differentiated instruction in classrooms. *Teaching and Teacher Education, 67*, 291-301.
- Takahashi, A., & McDougal, T. (2016). Collaborative lesson research: maximizing the impact of lesson study. *ZDM Mathematics Education, 48*(4), 513-526.
- Tomlinson, C. A. (2005). Grading and differentiation: Paradox or good practice? *Theory into Practice, 44*(3), 262-269.
- Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.
- UNESCO. (2009). *Policy guidelines on inclusion in education*. Paris, France: UNESCO.
- Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research, 49*(2), 127-152.
- Van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of pupil teachers: Calibration of an evaluation instrument and its value in predicting pupil academic engagement. *Studies in Educational Evaluation, 43*, 150-159.
- Van der Lans, R. M., Van de Grift, W. J. C. M., & Van Veen, K. (2017). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *The Journal of Experimental Education, 86*(2), 247-264.

- Van Halem, N., Goei, S. L., & Akkerman, S. F. (2016). Formative assessment in teacher talk during lesson studies. *International Journal for Lesson and Learning Studies*, 5(4), 313-328.
- Vesterinen, O., Toom, A., & Patrikainen, S. (2010). The stimulated recall method and ICTs in research on the reasoning of teachers. *International Journal of Research & Method in Education*, 33(2), 183-197.
- Vrikki, M., Warwick, P., Vermunt, J. D., Mercer, N., & Van Halem, N. (2017). Teacher learning in the context of Lesson Study: A video-based analysis of teacher discussions. *Teaching and Teacher Education*, 61, 211-224.
- Wan, S. W. Y. (2016). Differentiated instruction: Hong Kong prospective teachers' teaching efficacy and beliefs. *Teachers and Teaching*, 22(2), 148-176.
- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21, 343-356.
- Xu, H., & Pedder, D. (2015). Lesson Study: an international review of the research. In P. Dudley (Ed.), *Lesson Study. Professional learning for our time*. New York: Routledge.
- Ylonen, A., & Norwich, B. (2015). How Lesson Study helps teachers of pupils with specific needs or difficulties. In P. Dudley (Ed.), *Lesson Study. Professional learning for our time*. New York: Routledge.
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, pupil academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981-1015.

Tables and figures

Table 1. Sample descriptions ($N = 48$).

	Intervention group	Comparison group
Amount of teachers	26 (range: 1-5 teachers per school)	22 (range: 1-5 teachers per school)
Female %	53.8%	59.1%
Age (years)	$M = 44.2$, $SD = 13.1$ (range: 24-60)	$M = 43.3$, $SD = 13.3$ (range: 21-63)
Teaching experience (years)	$M = 18.2$, $SD = 11.7$ (range: 4-39)	$M = 14.3$, $SD = 10.6$ (range: 1-37)
Teacher qualification	M.Ed.: 42.3% B.Ed.: 50.0% In training: 7.7%	M.Ed.: 45.5% B.Ed.: 45.5% In training: 9.1%
Teaching subject	Languages: 50.0% Social sciences: 38.5% Sciences: 11.5%	Languages: 54.5% Social sciences: 40.9% Sciences: 4.5%

Notes: Teacher qualification ‘in training’ refers to teachers’ final stage of their B.Ed./M.Ed. teacher training programme. The subcategory ‘languages’ consists of Dutch, English, German, and French. The subcategory ‘social sciences’ consists of economics, history, geography and civics, and the subcategory ‘sciences’ consists of mathematics, physics, chemistry and biology.

Table 2. Composition of intervention group ($n = 26$)

School	Research project #	Amount of teachers		LS team composition	Use of case pupils	Time allocation (hours)	LS Facilitator
		<i>Male</i>	<i>Female</i>				
1	1	1	2	Content specific	Yes	27	External
2	1	4	0	Content specific	Yes	27	External
3	1	2	0	Content specific	Yes	27	External
4	1	0	5	Content specific	Yes	27	External
5	1	1	0	Content specific	Yes	27	External
6	2	0	4	Interdisciplinary	No	166	Internal
7	2	2	1	Interdisciplinary	No	166	Internal
8	3	2	2	Interdisciplinary	Yes	85	External & internal

Table 3. TSE mean scores and standard deviations in pretest and posttest for both groups.

	Pretest		Posttest		<i>N</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<i>Intervention group</i>					
Efficacy in pupil engagement	6.13	.64	6.30	.63	26
Efficacy in instructional strategies	6.30	.76	6.53*	.80	26
Efficacy in classroom management	7.06	.64	7.09	.69	26
<i>Comparison group</i>					
Efficacy in pupil engagement	6.16	.79	6.02	.62	22
Efficacy in instructional strategies	6.28	.82	6.40	.54	22
Efficacy in classroom management	7.00	.86	7.02	.87	22

Note: * $p < .05$ (two-tailed).

Table 4. ICALT mean scores and standard deviations in pretest and posttest for both groups.

Scale	Pretest		Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Intervention group</i>				
1. Safe and stimulating learning climate	3.54	.40	3.59	.44
2. Efficient classroom management	3.19	.57	3.38	.48
3. Clarity of instruction	2.95	.56	3.18*	.54
4. Activating learning	2.44	.52	2.71**	.59
5. Adaptive teaching	1.75	.70	2.04*	.83
6. Teaching learning strategies	1.81	.45	2.10*	.57
7. Pupil involvement	2.73	.70	2.88	.76
<i>Comparison group</i>				
1. Safe and stimulating learning climate	3.48	.52	3.49	.48
2. Efficient classroom management	3.40	.50	3.15*	.63
3. Clarity of instruction	3.10	.44	2.94	.69
4. Activating learning	2.45	.55	2.55	.56
5. Adaptive teaching	1.44	.37	1.69*	.49
6. Teaching learning strategies	1.70	.58	1.92	.62
7. Pupil involvement	2.85	.61	2.67	.78

Note: * $p < .05$ (two-tailed). ** $p < .01$ (two-tailed).

Table 5. Pearson correlations among TSE and ICALT subscales

<i>Subscales</i>	1	2	3	4	5	6	7	8	9	10
1. TSE Engagement	1									
2. TSE Instructional	.48**	1								
3. TSE Management	.56**	.48**	1							
4. ICALT Stimulating	.03	.01	.27	1						
5. ICALT Management	.20	.21	.43**	.56**	1					
6. ICALT Instruction	.27	.20	.48**	.72**	.78**	1				
7. ICALT Activating	.29*	.17	.31*	.67**	.60**	.85**	1			
8. ICALT Adaptive	.33*	.22	.30*	.38**	.45**	.53**	.70**	1		
9. ICALT Strategies	.14	-.07	.22	.53**	.35*	.66**	.63**	.32*	1	
10. ICALT Pupil	.06	.03	.32*	.69**	.69**	.74**	.74**	.46**	.60**	1

Note: The intercorrelations were measured using the posttest data. * $p < .05$; ** $p < .01$.

Figure 1. Research design.

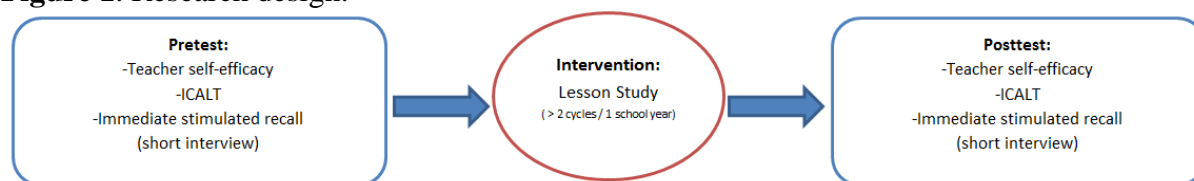
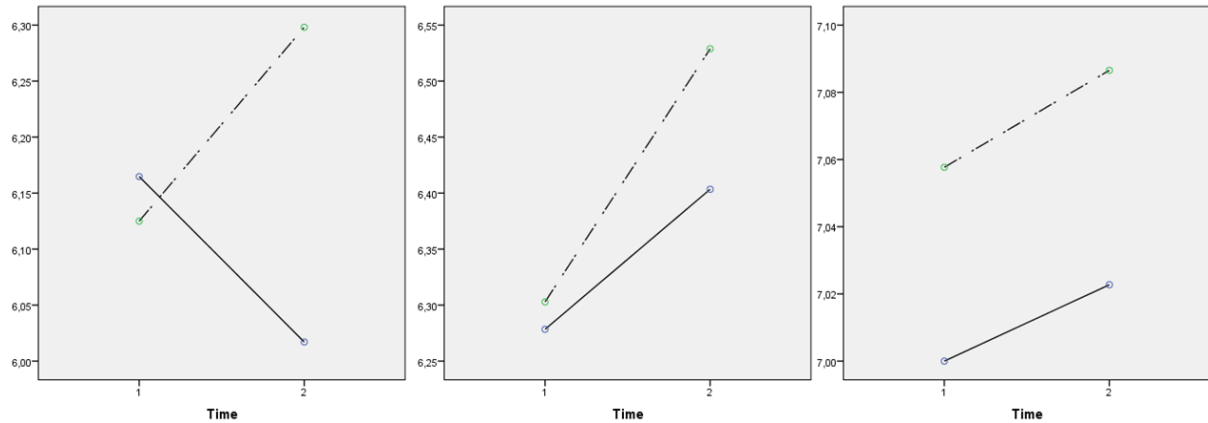
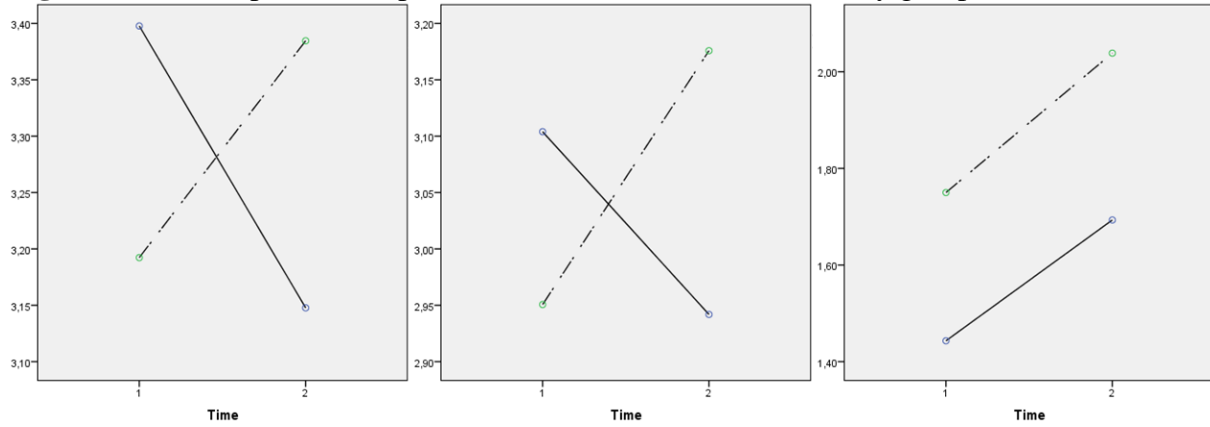


Figure 2. Teacher pretest and posttest levels of TSE by group.



Note: The X-axis illustrates pretest (1) and posttest (2) values. The Y-axis refers to the average values. The dashed lines connect the values of the pretest and posttest in the intervention group. The normal line illustrates this for the comparison group. The first, second and third graph relate respectively to the subscale efficacy in pupil engagement, efficacy in instructional strategies, and efficacy in classroom management.

Figure 3. Teacher pretest and posttest values of ICALT subscales by group.



Note: The X-axis illustrates pretest (1) and posttest (2) values. The Y-axis refers to the average values. The dashed lines connect the values of the pretest and posttest for the intervention group. The normal line illustrates this for the comparison group. The first, second and third graph relate respectively to the subscale (2) 'efficient classroom management', (3) 'clarity of instruction', and (5) 'adaptive teaching'.